

Universität Karlsruhe
Institut für Betriebs- und Dialogsysteme
Prof. Dr. Alfred Schmitt
Dipl.-Inform. B. Klimmek
Sommersemester 2005

Klausur Informatik IV

Mittwoch, 5. Oktober 2005

Name: _____
Vorname: _____

Matrikelnummer: _____

Beachten Sie:

- Trennen Sie vorsichtig die dreistellige Nummer von Ihrem Aufkleber ab. Sie sollten sie gut aufheben, um später Ihre Note zu erfahren.
- **Vor Beginn** der Klausur haben Sie 5 Minuten Zeit zum Lesen der Aufgabenstellungen.
- Schreiben Sie Ihre Lösungen auf die Aufgabenblätter. Benutzen Sie gegebenenfalls die Rückseiten. Notfalls können Sie weiteres Papier anfordern.
- Zum **Bestehen** der Klausur sind **20** der möglichen 60 **Punkte** hinreichend.
- Es sind **keine Hilfsmittel** zugelassen.
- Kleben Sie **nach der Klausur** den Aufkleber mit Ihrem Namen und Ihrer Matrikelnummer auf dieses Deckblatt. Schreiben Sie außerdem Namen und Matrikelnummer oben auf jedes bearbeitete Aufgabenblatt.

Aufgabe	1	2	3	4	5	6	7	8	Gesamt
Erreichte Punkte									
Mögliche Punkte	9	6	12	5	3	7	8	10	60

Aufgabe 1

Es werden k gleichverteilt-zufällige Einträge in eine anfangs leere Hash-Tabelle der Größe $n \geq 8$ geschrieben.

Mit linearer Suche wird sodann die erste nicht leergebliebene Hash-Zelle m gesucht.

Wie groß ist die Wahrscheinlichkeit dafür, daß $m = 8$ ist?

(Vereinfachen Sie ihr Ergebnis so weit wie möglich.)

$$\begin{aligned} P &= P(\text{Zellen Nr. } 1, \dots, 7 \text{ sind leer}) \cdot \\ &\quad P(\text{Zelle Nr. } 8 \text{ aber nicht leer}) \\ &= P(\text{Zellen Nr. } 1, \dots, 7 \text{ sind leer}) \cdot \\ &\quad [1 - P(\text{Zelle Nr. } 8 \text{ leer} \mid \text{Zellen } 1, \dots, 7 \text{ leer})] \\ &= \left(\frac{n-7}{n}\right)^k \cdot \left[1 - \left(\frac{n-8}{n-7}\right)^k\right] \end{aligned}$$

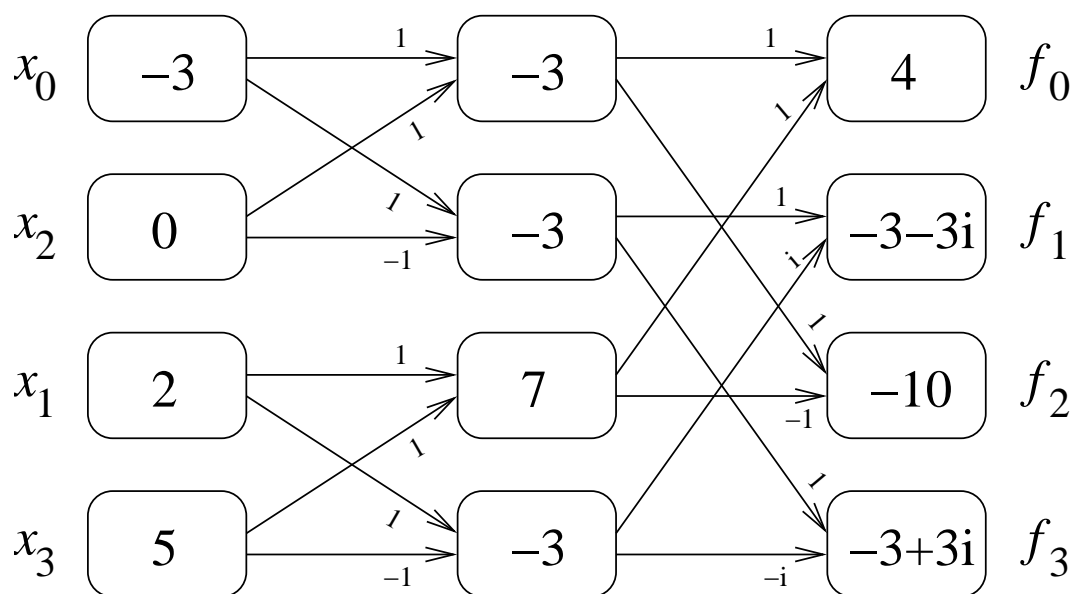
Die Lösung läßt sich, in anderer Form, auch über Indexfolgen (wie in der Vorlesung) erhalten:

$$\begin{aligned} P &= \frac{|\text{günstige Indexfolgen}|}{|\text{mögliche Indexfolgen}|} \\ &= \frac{\sum_{i=1}^k \binom{k}{i} (n-8)^{k-i}}{n^k} \end{aligned}$$

Dabei ist i die Anzahl ($0 < i \leq k$) von Einträgen in der 8-ten Zelle.

Aufgabe 2

Führen Sie für das Quadrupel $X = (x_0, x_1, x_2, x_3) = (-3, 2, 0, 5)$ eine Schnelle Fouriertransformation (FFT) mithilfe des Butterfly-Schemas durch und geben Sie das Ergebnis-Quadrupel $f(X) = (f_0, f_1, f_2, f_3)$ an.



Also ist für $\omega = i$ das Ergebnis:

$$f(X) = (4, -3 - 3i, -10, -3 + 3i)$$

(Mit $\omega = -i$ erhält man die konjugierten Werte $f(X) = (4, -3 + 3i, -10, -3 - 3i)$, was auch nicht falsch ist.)

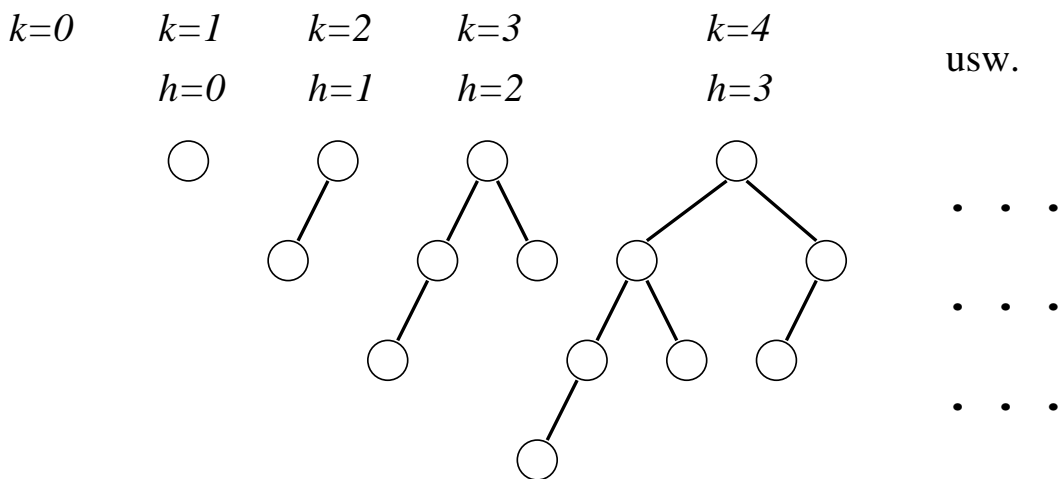
Aufgabe 3

Betrachten Sie einen maximal unausgeglichene Binärbaum der Höhe h , der gerade noch die AVL-Balancebedingung erfüllt. Jedes Blatt hat, von der Wurzel aus gerechnet, eine bestimmte Tiefe. Um wie viele Tiefenniveaus $d = d(h)$ kann die Lage der Blätter innerhalb eines solchen Baums variieren?

(Tip: Wie kann ein solcher (ohne Beschränkung der Allgemeinheit: maximal linkslastiger) Baum induktiv/rekursiv aufgebaut werden?)

Es ist kein strenger Beweis erforderlich, aber eine gute Beweisskizze bzw. -idee.

Ein maximal linkslastiger, aber noch die AVL-Bedingung erfüllender Binärbaum kann konstruiert werden, indem (für $k > 1$) Baum Nr. k aus einem Wurzelknoten besteht, an den links der Baum Nr. $k-1$ als Unterbaum gehängt wird und rechts der Baum Nr. $k-2$. (Basisfälle: der Baum Nr. 1 besteht nur aus einem Knoten, Baum Nr. 0 ist leer.)



Der Baum hat dann die Höhe $h = k - 1$; und er ist balanciert, denn die Höhe der beiden Unterbäume differiert genau um 1. (Übrigens ist dann das ganz linksaußen gelegene Blatt im Baum auf

unterstem Niveau, das am weitesten rechtaußen gelegene Blatt dasjenige auf oberstem Niveau.)

Man erhält für die höchst- und tiefstgelegenen Blätter abhängig von h die folgenden Werte:

Baum der Höhe $h =$	0	1	2	3	4	5	6	...
oberstes Blatt-Niveau $o =$	0	1	1	2	2	3	3	...
unterstes Blatt-Niveau $u =$	0	1	2	3	4	5	6	...
Differenz $d = u - o =$	0	0	1	1	2	2	3	...

Denn es gilt ja aufgrund der Struktur dieses Baumes $u(h) = h$ und $o(h) = o(h - 2) + 1$ (also offenbar $o(h) = \lceil h/2 \rceil$, vgl. Tabelle).

Damit ist der Abstand zwischen höchst- und tiefstgelegenen Blatt in einem solchen AVL-Baum der Höhe h offensichtlich der Wert

$$d(h) = \lceil h/2 \rceil$$

Aufgabe 4

Betrachten Sie das *travelling salesperson problem* (TSP) mit $n = 5000$ Punkten $\mathbf{p}_i = (x_i, y_i)$ in der Ebene.

- a) Warum werden zur (approximativen) Lösung des TSP heuristische Verfahren eingesetzt?

Das TSP gehört zur Klasse der NP-vollständigen Probleme und erfordert (nach dem derzeitigen Kenntnisstand) exponentielle Laufzeit; es gibt $n! = 5000! \approx e^{5000 \cdot \log 5000} > e^{40000} \approx 10^{17000}$ potentielle Lösungen. Diese alle auszutesten, ist ein hoffnungslos langatmiges Unterfangen.

Daher gibt man sich mit – wenn nicht optimalen, so doch – “guten” Lösungen zufrieden.

(1 Punkt)

- b) Kann im Rahmen einer solchen Heuristik das Gradientenabstiegs-Verfahren angewendet werden?

Nein.

(4 Punkte)

- Wenn ja: Geben Sie schemenhaft eine mathematisch-algorithmische Spezifikation des Gradientenabstiegs an.

- Wenn nein: Warum nicht?

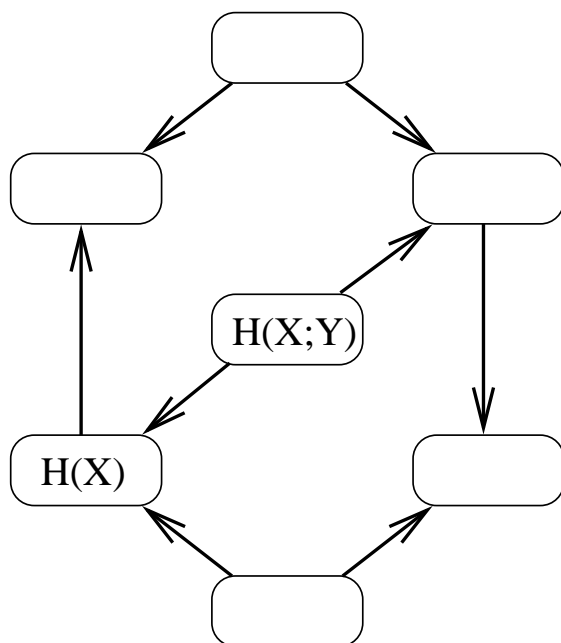
Weil der Lösungsraum nicht stetig ist, sondern diskret. *Das Gradientenverfahren setzt die Differenzierbarkeit der Bewertungsfunktion voraus.* Die \mathbf{p}_i sind beim TSP gegeben (und fest), gesucht ist eine von $n!$ Permutationen dieser Punkte.

5 Punkte

Aufgabe 5

Ergänzen Sie das folgende Diagramm. Es fehlen noch die folgenden Kenngrößen zur Beschreibung eines Übertragungskanals: Zielinformation $H(Y)$, Fehlinformation $H(Y|X)$, Äquivokation $H(X|Y)$, Gesamtinformation $H(X, Y)$.

Überlegen Sie zunächst, was die Pfeile symbolisieren könnten.

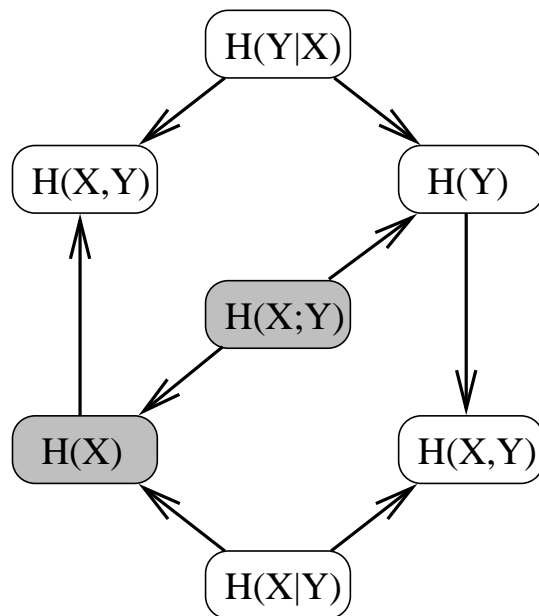
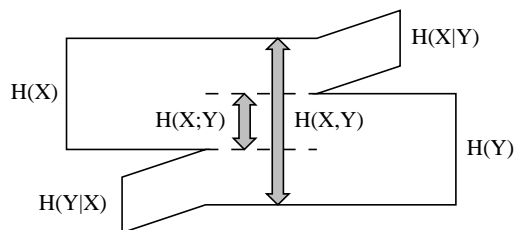


Weil jeweils mehrere Pfeile auf einen Knoten zeigen (und außerdem $H(X; Y)$ ein Summand ist von $H(X) = H(X; Y) + H(X|Y)$), liegt es nahe, die Pfeil-Relation “ \rightarrow ” als “ist Summand von” zu übersetzen.

Dann ergibt sich, angesichts der Beziehungen

$$H(X) - H(X|Y) = H(X; Y) = H(Y) - H(Y|X) \text{ und}$$

$$H(X) + H(Y|X) = H(X, Y) = H(Y) + H(X|Y) :$$



(Weil für alle Terme $H(\dots)$ gilt, daß sie nicht-negativ sind, kann die Pfeil-Beziehung “ \rightarrow ” auch als “ \leq ”-Relation gedeutet werden, was zum selben Ergebnis führt.)

Aufgabe 6

- a) Führen Sie eine Burrows-Wheeler-Transformation auf der Zeichenkette $w = \text{BLUBBLUB}$ durch. Geben Sie als Ergebnis alle Informationen an, die die Rücktransformation benötigt, um w wieder herzustellen.

$$M = \begin{array}{cccccccc} B & L & U & B & B & L & U & B & (3) \\ L & U & B & B & L & U & B & B & (5) \\ U & B & B & L & U & B & B & L & (7) \\ B & B & L & U & B & B & L & U & (1) \\ B & L & U & B & B & L & U & B & (4) \\ L & U & B & B & L & U & B & B & (6) \\ U & B & B & L & U & B & B & L & (8) \\ B & B & L & U & B & B & L & U & (2) \end{array}$$

$$M' = \begin{array}{cccccccc} B & B & L & U & B & B & L & U & (1) \\ B & B & L & U & B & B & L & U & (2) \\ B & L & U & B & B & L & U & B & (3) \leftarrow \text{Original} \\ B & L & U & B & B & L & U & B & (4) \leftarrow \text{Original} \\ L & U & B & B & L & U & B & B & (5) \\ L & U & B & B & L & U & B & B & (6) \\ U & B & B & L & U & B & B & L & (7) \\ U & B & B & L & U & B & B & L & (8) \end{array}$$

Dann ist die transformierte Zeichenkette v identisch mit der letzten Spalte $L = \text{UUBBBLL}$ und der Startindex ist $k = 3$ oder $k = 4$ (egal).

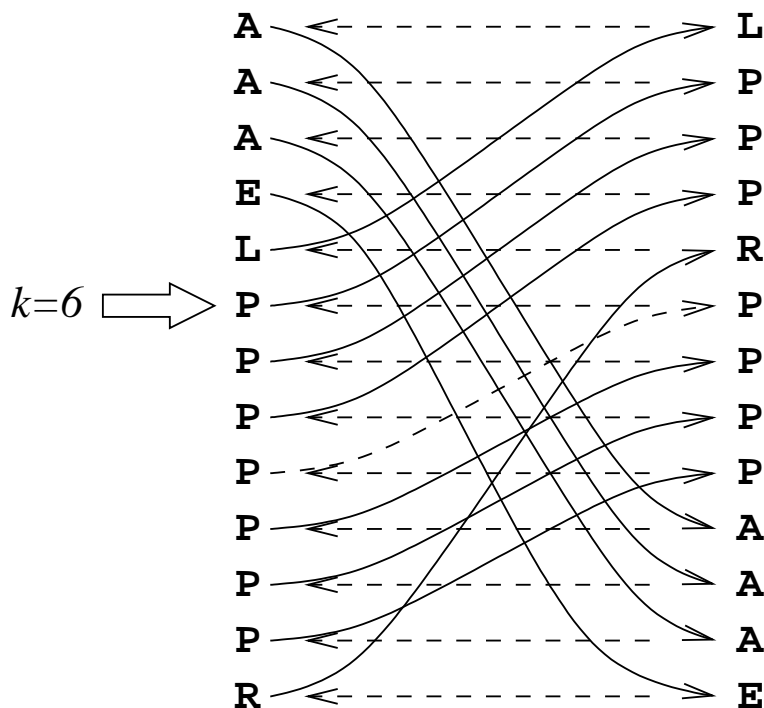
(3 Punkte)

- b) Dekodieren Sie die lauffängenkodierte Folge $(L, 1), (P, 3), (R, 1), (P, 4), (A, 3), (E, 1)$, führen Sie für die Ergebnis-Zeichenkette v die inverse Burrows-Wheeler-Transformation (Dekodierung) mit Startindex $k = 6$ durch und geben Sie das Ergebnis w an.

Nach Lauffängen-Dekodierung ist die Zeichenkette $v = \mathbf{LPPPRPPPPAAAE}$.

Diese ist nun die letzte Spalte L der Matrix M' . Die erste Spalte F ist L in sortierter Reihenfolge: $F = \mathbf{AAAE L P P P P P P P R}$.

Nun die Burrows-Wheeler-Dekodierung:



Die inverse Transformation ergibt demnach die Zeichenkette $w = \mathbf{PAPPERLAPPAPP}$.

(4 Punkte)

7 Punkte

Aufgabe 7

Über dem Alphabet $X = \{A, B, C\}$ sei eine Zeichenkette w gegeben durch die LZW-Kodierung

0, 1, 2, 3, 5, 4, 6.

Dabei seien die Wörterbuch-Einträge anfänglich gegeben durch: $0 \equiv A$, $1 \equiv B$, $2 \equiv C$.

- a) Wie lang ist die Zeichenkette w ?

LZW-Dekodierung:

Eingabe	Ausgabe	Wörterbuch
0	A	3=AB
1	B	4=BC
2	C	5=CA
3	AB	6=ABC
5	CA	7=CAB
4	BC	8=BCA
6	ABC	

Die Zeichenkette $w = ABCABCABCABC$ hat also 12 Zeichen.

(3 Punkte)

- b) Um wie viele Bits differiert die Länge der LZW-Kodierung von w von der Länge einer Huffman-Kodierung von w ?

Huffman-Kodierung:

Da in w alle drei Zeichen gleich häufig auftreten, ist es egal, welches das “privilegierte” 1-Bit-Kodewort zugeordnet bekommt; die beiden restlichen Zeichen benötigen dann je 2 Bit Kodelänge.

Daher ergibt sich

$$\begin{aligned} \text{code}(w) &= \text{code}((ABC)^4) \\ &= 4 \cdot (1 \text{ bit} + 2 \text{ bit} + 2 \text{ bit}) = 20 \text{ bit} \end{aligned}$$

LZW-Kodierung:

Da die sieben Codes im sich im Zahlenraum $\{0, \dots, 6\}$ bewegen (es ist $|\{0, \dots, 6\}| \leq 8 = 2^3$), genügt es, sie mit je 3 Bit binär zu kodieren. Gesamtkodelänge wäre dann $7 \cdot 3 \text{ bit} = 21 \text{ bit}$.

Die Differenz beträgt also $|20 - 21| = 1 \text{ Bit}$.

(5 Punkte)

8 Punkte

Aufgabe 8

Kreuzen Sie für folgende Aussagen an, ob sie wahr oder falsch sind.

(Für jede richtige Antwort gibt es einen Punkt, für jede falsche Antwort wird ein Punkt abgezogen. Die gesamte Aufgabe wird mit einer nichtnegativen Punktezahl gewertet.)

<i>Aussage</i>	wahr	falsch
Der Schnitt zweier rationaler Dünns-Mengen ist stets in einer rationalen Dünns-Menge enthalten.	wahr	
Die Zahl der im Mittel erforderlichen Schlüsselvergleiche “<” zur Elementsuche in einem B-Baum der Ordnung k weicht ca. um den Faktor $k/2$ von der in einem AVL-Baum (mit denselben Schlüsseln) ab.		falsch
Die diskrete Fouriertransformation ist verlustfrei.	wahr	
Es ist stets $H(X) \geq H(X Y)$.	wahr	
Der O-Kalkül verhält sich neutral gegenüber multiplikativen Konstanten $C > 0$, nicht aber gegenüber additiven Konstanten D .		falsch
Eine dreistufige FFT kann statt mit Potenzen von $\omega = e^{2\pi i/8}$ auch mit Potenzen von $\omega = e^{2\pi i \cdot 2/8}$ realisiert werden.		falsch
Nach der Burrows-Wheeler-Transformation $w \mapsto v$ gilt für die Shannon-Informationsgehalte der beiden Zeichenketten: $H(v) < H(w)$.		falsch
Die Burrows-Wheeler-Transformation einer Zeichenkette der Länge n kann mit Speicheraufwand von $O(n)$ implementiert werden.	wahr	
Es gibt Elementaroperationen auf AVL-Bäumen, bei denen eine (Doppel-)Rotation nicht ausreicht, um die AVL-Eigenschaft wieder herzustellen.	wahr	
Jeder optimale Kode variabler Länge, der eindeutig entzifferbar ist, ist auch sofort entzifferbar.		falsch

10 Punkte